# REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188



**1. AGENCY USE ONLY** *(Leave blank)*

**2. REPORT DATE** 3/15/96

**3. REPORT TYPE AND DATES COVERED** Final Technical 6/1/94 - 1/31/96

**4. TITLE AND SUBTITLE**
Further Research on Super Auditory Localization for Improved Human-Machine Interfaces

**5. FUNDING NUMBERS**
F49620-94-1-0236
2313/~~CS~~ BS

**6. AUTHOR(S)**
Nathaniel Durlach

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Research Laboratory of Electronics
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

AFOSR-TR-96
0154

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Air Force Office of Scientific Research/NL
110 Duncan Avenue, Suite B115
Bolling A.F.B., DC 20332-0001

**10. SPONSORING/MONITORING**

19960404 072

**11. SUPPLEMENTARY NOTES**
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**
Approved for public release; distribution unlimited.

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*

The general objectives of our initial work on Super Auditory Localization were to "determine, understand, and model the perceptual effects of altered localization cues." We had initially intended to conduct this work using a virtual-environment (VE) system for visual as well as auditory stimulation, and to include examination of a wide variety of transformations (rotations, scalings, filterings, asymmetries, exponentiations). As will be seen in the following discussion, we have made substantial progress towards our general objectives. However, our work was conducted using a hybrid VE in which the acoustical stimulation was virtual but the visual stimulation was real, we focused on only one family of azimuthal transformations, and we made no effort to measure our own HRTFs. The decision to use available HRTFs rather than to construct our own was based on the realization that, at least for our purposes, such work would have a relatively low payoff-to-effort ratio compared to other work that needed to be done. Both the hybrid VE and the azimuthal transformation used are described in Sec. I.B below.

The gaps between our stated objectives and our actual accomplishments are the result of a number of factors. The first and most important is that the total funding we have received constitutes only a small fraction of the funding that we requested in order to achieve the above-stated goals. Whereas our proposal totalled roughly $1,501,000, the total amount of funds that we have actually received to date for this project is roughly $700,000 ($650,000 from AFOSR and $50,000 from NASA). (All figures are Total Costs, not Direct Costs). Secondary factors include (1) the complexity of the subject addressed, (2) the relatively high cost and limited performance of the VE equipment that was available during the working period of this grant, and (3) the departure from MIT of a key research scientist assigned to this research (X. D. Pang, for personal reasons). In light of these factors, we believe that our progress, discussed in detail in the following subsections, has been substantial.

**14. SUBJECT TERMS**

**15. NUMBER OF PAGES**

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | UNCLASSIFIED | UL |

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

# Further Research on Super Auditory Localization for Improved Human-Machine Interfaces

**Grant F49620-94-1-0236**

# Final Report

**Principal Investigator:** Nathaniel Durlach
Massachusetts Institute of Technology
36-709
77 Massachusetts Ave.
Cambridge, MA 02139

# Summary

The general objectives of our initial work on Super Auditory Localization were to "determine, understand, and model the perceptual effects of altered localization cues." We had initially intended to conduct this work using a virtual-environment (VE) system for visual as well as auditory stimulation, and to include examination of a wide variety of transformations (rotations, scalings, filterings, asymmetries, exponentiations). As will be seen in the following discussion, we have made substantial progress towards our general objectives. However, our work was conducted using a hybrid VE in which the acoustical stimulation was virtual but the visual stimulation was real, we focused on only one family of azimuthal transformations, and we made no effort to measure our own HRTFs. The decision to use available HRTFs rather than to construct our own was based on the realization that, at least for our purposes, such work would have a relatively low payoff-to-effort ratio compared to other work that needed to be done. Both the hybrid VE and the azimuthal transformation used are described in Sec. I.B below.

The gaps between our stated objectives and our actual accomplishments are the result of a number of factors. The first and most important is that the total funding we have received constitutes only a small fraction of the funding that we requested in order to achieve the above-stated goals. Whereas our proposal totalled roughly $1,501,000, the total amount of funds that we have actually received to date for this project is roughly $700,000 ($650,000 from AFOSR and $50,000 from NASA). (All figures are Total Costs, not Direct Costs). Secondary factors include (1) the complexity of the subject addressed, (2) the relatively high cost and limited performance of the VE equipment that was available during the working period of this grant, and (3) the departure from MIT of a key research scientist assigned to this research (X. D. Pang, for personal reasons). In light of these factors, we believe that our progress, discussed in detail in the following subsections, has been substantial.

# I. Accomplishments/New Findings

## I.A. Equipment Issues

The work originally envisioned depended strongly on the availability of adequate technology for the presentation and control of acoustic and visual stimuli. Because the technology available proved to be less than adequate, a number of our research goals were scaled back or altered to fit the capabilities of the devices available. In addition, the development of improved equipment became a goal of the project.

One of the original objectives of the project was to investigate the use of auditory localization cues that exceeded the range of normal cues (e.g., interaural time differences that exceeded those that occur naturally). Unfortunately, the Convolvotron (the special-purpose auditory spatialization system used to synthesize localization cues in our experiments) was designed to present normal localization cues and was found to be incapable of presenting localization cues outside the normal range. Although the hardware in the Convolvotron is capable of generating abnormally large interaural time differences for a single source in real-time, it cannot do so for four sources

simultaneously. Even making use of the Convolvotron for a single source with abnormally large interaural differences proved impossible due to software constraints. Furthermore, the Convolvotron can store HRTFs for only a small number of source positions and performs a spectral interpolation to simulate source positions between these stored locations. Although the possibility of significant interpolation error was a troubling (but unavoidable) problem even for the use of normal HRTFs, the errors introduced by interpolation of super-normal HRTFs would be even larger. Consequently, HRTFs containing larger-than-normal localization cues were not used with the existing Convolvotron. Because of these limitations with the Convolvotron, the acoustical localization cues for the reported experiments were drawn from the pool of normal acoustical cues (which were stored in the Convolvotron), and cue alterations were achieved by changing the mapping between these cues and the direction of the source relative to the head.

In addition to restricting the magnitude of localization cues simulated, the Convolvotron /head-tracking acoustic VE suffered from time delays and other distortions (e.g., the distortions induced by spatial interpolation of HRTFs discussed above). A great deal of this delay was the result of the Bird tracker employed. This tracker, although state-of-the-art when purchased, suffers from time delays on the order of tens of milliseconds. In order to test the importance of these effects, alternate synthesis methods were developed.

Fig. 1. The M.I.T. pseudophone configured to present supernormal interaural delays.

A second acoustic synthesis device, based on the system described in Loomis, Hebert, and Cincinelli (1990), was procured in order to further test the effects of artifacts associated with the use of the Convolvotron (as well as to explore the use of simplified cues). This second system uses highly simplified interaural level and monaural spectral cues. Since it is an analog system, interpolation of cues is not necessary with this device, as it is with the Convolvotron.

A pseudophone was designed and built at M.I.T. which employs microphones that can be located at various points relative to the head and that are connected to headphones worn by the subject. The pseudophone will allow presentation of unnaturally large interaural differences (in amplitude as well as time, although Fig. 1 shows the system configured solely to increase interaural time delay) which are perfectly correlated with the wearer's movements with essentially no delay between head movement and change in stimulus characteristics. Also, background sounds will go through the same transformation as the intended targets since the auditory rearrangement depends upon the physical geometry of the microphones rather than synthesis of acoustic cues by signal processing methods. Attenuation of natural, unprocessed sounds is achieved by the use of insert earphones and acoustic muffs.

The time delays and small working volume associated with existing tracking systems inspired the development of an inertial tracking system in addition to development of the pseudophone. This tracker[1] will provide a large working volume, increased resolution, and better dynamic performance than existing tracking devices. A prototype inertial tracker for the three degrees of freedom associated with head orientation has been tested and will be ready for use in the near future.

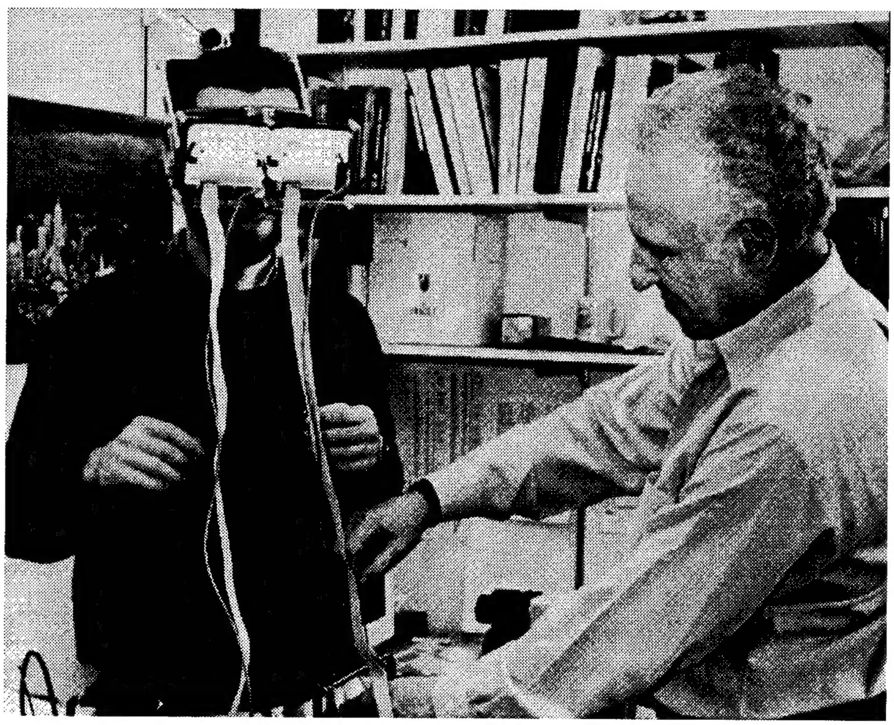

Fig. 2. The M.I.T. head-mounted display (HMD).

Development of a visual virtual environment (VE) was originally undertaken to provide more flexible control of visual stimuli in the current project. A stereo head-mounted display (HMD) was developed in-house (see Fig. 2). This system, built from commercially available components, proved compact, lightweight, and portable. The HMD is completely untethered so that subjects

[1]Development of the inertial tracker has been partially supported by NASA.

can walk around freely when wearing it. The task of integrating the HMD with a graphics machine and the existing auditory VE in order to synthesize visual stimuli proved to be much more costly in both time and effort than was originally anticipated. While some progress on the development of a visual VE was made in the first year of the project, these efforts were put off in order to concentrate more fully on adaptation experiments that could be performed with the hybrid environment.

Since our last report, we have also obtained a display device made by Tucker-Davis Technologies (under a separate Navy contract) which can process stimuli with larger-than-normal interaural differences. We are in the process of developing software which simulates larger-than-normal HRTF cues with this device. We have also acquired a Crystal River Engineering Gargantutron, providing us with additional hardware on which to test and develop experiments, and a FASTRAK headtracker, allowing us to develop a system with a shorter latency than was possible with the Bird tracker.

## I.B. Experimental Work

Adaptation to altered auditory localization cues was investigated by presenting simulated acoustic cues and real visual cues. Acoustic sources were "spatialized" by the Convolvotron, the special-purpose signal-processing system made by Crystal River Engineering and discussed above. The Convolvotron takes as inputs the source signal to be spatialized and the instantaneous position of the source relative to the listener's head and generates the binaural signals appropriate for a source from the specified position. In our system, the relative source position was calculated by a PC from the absolute position of the source to be simulated and the instantaneous orientation of the listener's head (reported to the PC by the Bird, a commercial head-tracking system).

This auditory virtual environment was used to simulate sources from one of thirteen positions around the listener at 0 degrees elevation, from -60 to +60 degrees in azimuth. These positions were indicated visually by a 3-foot-diameter arc of lights, which were clearly labelled (1 to 13) from left to right. These lights constituted our "real" visual display and were used to present visual spatial information about the simulated auditory sources presented to our subjects.

Auditory localization cues were transformed in this project by remapping the relationship between source position and Head Related Transfer Functions (or HRTFs) as discussed in Sec. B-4-a-iii. This approach made optimal use of the Convolvotron and at the same time ensured (to the extent possible) that a given source position was perceived as a compact image. The mapping function used is given by Eq. B-10, repeated here for convenience:

$$f_n(\theta) = \frac{1}{2}\tan^{-1}\left[\frac{2n\sin(2\theta)}{1-n^2+\left(1+n^2\right)\cos(2\theta)}\right]. \tag{1}$$

This mapping is shown in Fig. B-7 for different values of n. For values of $n > 1$, source positions are displaced laterally relative to normal cues. The differences in localization cues for two sources in the frontal region (from -30 to +30 degrees in azimuth) are larger than normal with this remapping, while two locations off to the side give rise to more similar cues than are normally

heard. With such a transformation, subjects were expected to show better than normal resolution in the front and reduced resolution on the side, creating an enhanced "acoustic fovea" in which super auditory localization could occur. In addition to affecting resolution, however, this type of transformation was also expected to cause a bias whereby sources were perceived farther off-center than were their actual locations. The main questions of the study were whether (1) bias could be overcome by subjects over time, so that they interpreted the new acoustic mapping of source position accurately, and (2) resolution was enhanced as expected in the "acoustic fovea". In all of our experimental work to date, attention has been focused on the identification of source azimuth.

### I.B.1. Experiment A

The basic experimental protocol consisted of a sequence of interleaved training and test runs. Each test run in the sequence consisted of 26 trials of a 13-alternative angle identification experiment. Test stimuli consisted of a 500 ms long click-train from one of 13 azimuthal positions separated by 10 degrees (ranging from -60 to +60 degrees). These positions corresponded to the positions of the lights, which were clearly numbered from left to right in an arc around the subject. Subjects had to face forward during each test stimulus or the trial was discarded. No correct-answer feedback was given and the lights were not used during the test runs. After each source was presented, the subject entered the number of the source position on a laptop keyboard.

During training runs, the subject was asked to track the source (whose position was chosen randomly for each trial from the set of 13 positions) by turning to point his/her nose to the correct location. During training, the light at the simulated acoustic location was turned on simultaneously with the acoustic source. In this manner, the subject became familiar with the mapping between source position, acoustic cues, and head orientation.

Each session (which lasted roughly 1.5 hrs) in this basic protocol consisted of the following sequence of test and training runs:

| | |
|---|---|
| Test using normal cues | (1n) |
| Train using normal cues | |
| Test using normal cues | (2n) |
| - 5 minute break - | |
| Test using altered cues | (1a) |
| Train using altered cues | |
| Test using altered cues | (2a) |
| Train using altered cues | |
| Test using altered cues | (3a) |
| Train using altered cues | |
| Test using altered cues | (4a) |
| - 5 minute break - | |
| Test using altered cues | (5a) |
| Train using altered cues | |
| Test using normal cues | (3n) |
| Train using normal cues | |
| Test using normal cues | (4n) |
| Train using normal cues | |
| Test using normal cues | (5n) |

Test Runs 1n, 1a, 5a, and 3n were analyzed in order to investigate how performance changed over the course of each session. Run 1n provided a control against which other runs could be compared. Run 1a provided a measure of the immediate effect of the transformed cues. Any decrease in effect was found by comparing Runs 1a and 5a. Finally, Run 3n showed any negative after-effects due to exposure to the altered cues. The training and testing runs performed after Run 3n were included in order to help the subject re-adapt to normal cues. No special attention was given in these preliminary experiments to the issues of conditional or dual adaptation (e.g., Welch, 1978; Welch et al., 1993).

Using this paradigm, each of four subjects completed 8 identical sessions. Performance did not change significantly from the first to final session.

A couple of different data processing schemes were investigated. In one method based on standard psychophysical analysis methods (e.g. Durlach and Braida, 1972), the confusion matrix (matrix whose entry i, j corresponded to the number of responses i given when position j was presented) was analyzed for each subject and run, with multiple sessions combined within each such matrix (on the whole, we found comparatively little variation across sessions). With this approach, each source presentation was assumed to result in a stochastic decision variable with a Gaussian distribution along some internal decision axis. The mean of the distribution was assumed to depend monotonically on the source position and the variance was assumed equal for all source positions. Further, the decision axis was assumed to be broken into 13 contiguous regions corresponding to the 13 possible responses. In this model, if the sample of the decision variable fell into region 'i', the subject would respond "i". With these assumptions, a gradient-descent numerical algorithm was implemented to find the estimates of means and variances that maximized the likelihood of observing the given confusion matrix. From these maximum likelihood estimates, the sensitivity $d_i'$ (a measure of the ability of the subject to discriminate between source positions i and i + 1) and bias $\beta_i$ (a measure of the perceptual bias when position i is presented) were derived. While theoretically elegant, the solutions found with this method proved to be overly sensitive to outliers in the responses and numerically unstable.

In the second method, which proved to be both simpler and more robust, the average response and the standard deviation in response was found for each of the 13 possible locations for Runs 1n, 1a, 5a, and 3n (averaged across 8 sessions) for each subject. These two statistics (average response and standard deviation in response) were then used to estimate both resolution and bias for each run during the course of a session. Resolution between adjacent pairs of positions was estimated as the difference in mean responses normalized by the average of the standard deviations for the two positions. Bias (which is traditionally used to measure adaptation) was estimated as the difference between mean response and correct response, normalized by the standard deviation for the position. These metrics were averaged across subjects to generate a concise summary of results for each run (as with the variation across sessions, the variation across subjects was found to be relatively modest). This second approach determines estimates of $d_i'$ and $\beta_i$ that approach the maximum likelihood estimates found in processing method 1 as the number of response

categories increases.[2]

Fig. 3 shows Experiment A bias results for runs 1n, 1a, 5a, and 3n as a function of source position (In this figure, as well as those that follow, the index i in $d_i'$ and $\beta_i$ has been omitted for simplicity). Normal-cue runs (1n and 3n) are plotted with circles; altered-cue runs (1a and 5a) with squares. The open symbols represent runs prior to altered cue training exposure (1n and 1a) while filled symbols correspond to the "adapted" results (5a and 3n). Results from Run 1n (open circles) showed some systematic biases, although these errors were significantly smaller than those found in other runs.

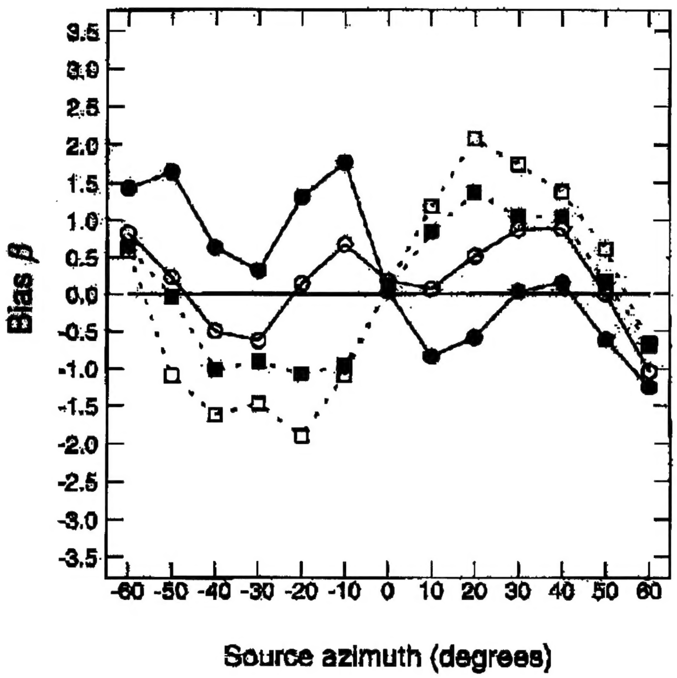


Fig. 3.Bias results for Experiment A. Normal cue tests are shown with circles, altered cue tests with squares. Open symbols represent tests prior to altered-cue exposure, filled symbols tests after exposure. The index i in $d_i'$ and $\beta_i$ has been omitted for simplicity.

In all bias results, there was an edge effect due to the experimental paradigm: since responses were limited to the 13 positions used, bias had to be positive (or zero) for the leftmost position (at -60 degrees azimuth) and negative (or zero) for the rightmost position (at +60 degrees azimuth). A strong bias occurred in Run 1a (open squares) in the direction predicted by the transformation and the aforementioned edge effect (subjects heard sources farther off-center than they were except for the leftmost and rightmost positions). Results from Run 5a (filled squares) showed a clear reduction in bias over the whole range of positions tested; however, this adaptation was not complete. Bias was reduced by roughly 30 percent with this experimental protocol. Finally, a negative after-effect is seen in the results from Run 3n (filled circles), where a strong bias was

[2]For experiments C and D, which used a pointing rather than identification response method, this second processing scheme yields the maximum likelihood estimates of $d_i'$ and $\beta_i$.

found in the direction opposite that induced by the altered cues.

Resolution results from Experiment A are shown in Fig. 4. Resolution for normal cue runs showed a systematic pattern (which may be due to systematic dependencies of the accuracy of the simulation on source position) which was consistent for pre- and post-exposure runs. Of more interest is the comparison between normal- and altered-cue results. As expected with the transformation employed, resolution was enhanced for positions in the central region and decreased at the edges of the range.

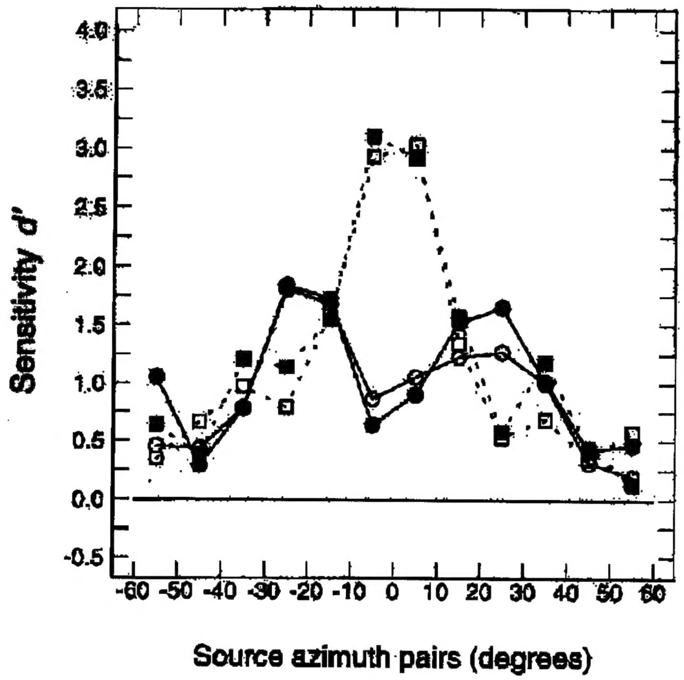


Fig. 4. Resolution results for Experiment A. See Fig. 3 caption.

### I.B.2. Experiment B

Since only partial adaptation was found with the basic paradigm, a minor alteration in the stimuli was made to try to get more complete adaptation. Experiment B was identical to experiment A, except that a more complete "acoustic field" (analogous to the visual field discussed in Radeau and Bertelson, 1976) was simulated. Along with the click-train target, continuous sources were simulated outside of the range of target positions: a music source (Handel, 1740) from -90 degrees, and a voice (Auel, 1980) from 180 degrees. Since both -90 and 180 degrees are mapped to the same position with the remapping function $f_3(\theta)$ (see Eq. B-10), these "stable" sources were presented from roughly the same positions during both normal- and altered-cue runs. During training runs, the expectation that each source remained in one exo-centric position as subjects turned their heads provided additional information about the transformation. These sources were added in an effort to make the acoustic field more complex and rich in information, since some studies of visual or auditory illusory effects show a dependence on the number of sources visible or audible (e.g., Lackner, 1983).

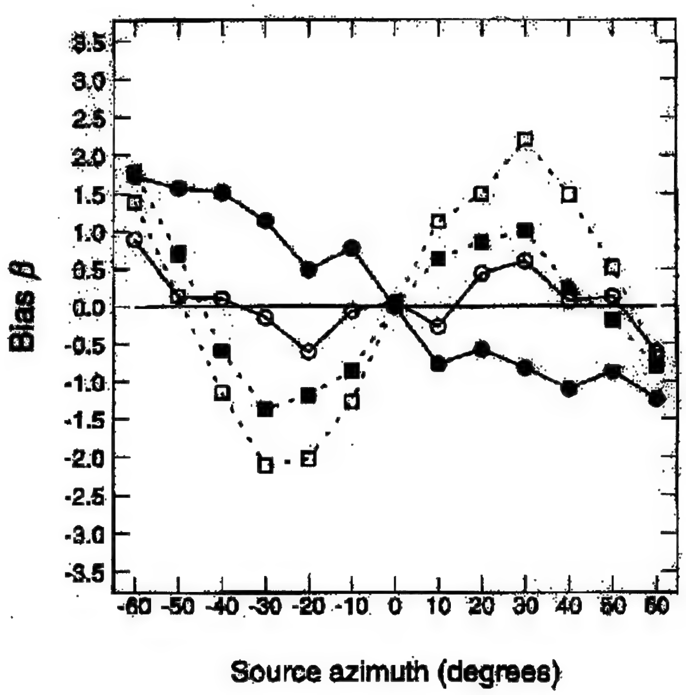


Fig. 5. Bias results for Experiment B. See Fig. 3 caption.

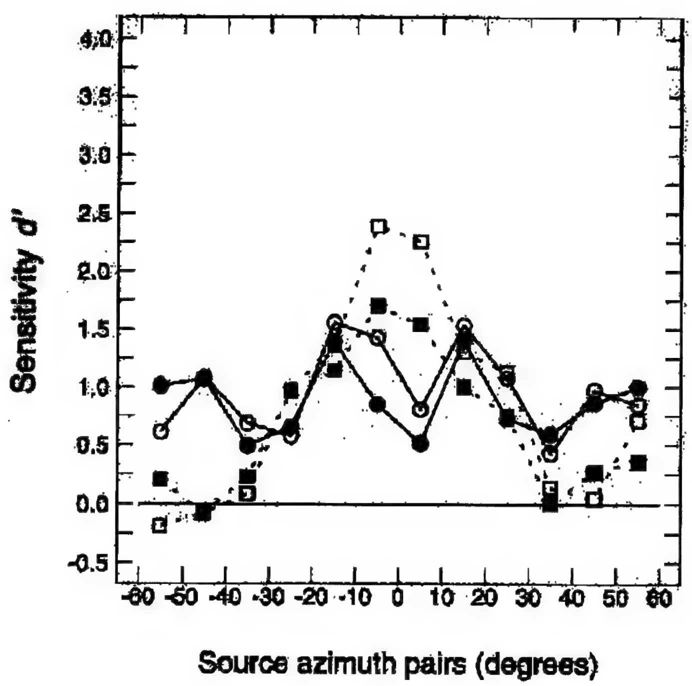


Fig. 6. Resolution results for Experiment B. See Fig. 3 caption.

Eight subjects performed Experiment B. Analysis yielded the bias results shown in Fig. 5 and the resolution results in Fig. 6. Bias results were very similar to those of Experiment A, with a

strong immediate effect, a reduction of roughly 30 - 50 percent with exposure, and a strong negative after-effect.

The resolution results for normal cue runs in Experiment B showed the same systematic variation as those of Experiment A. Resolution for the first altered cue run in Experiment B was similar to that of the first experiment, although the increase in resolution for the center two pairs of positions was somewhat smaller than that seen in Experiment A. Of more interest, however, were the resolution results for the final altered cue run. In Experiment B, resolution appeared to decrease significantly for the center positions with exposure to the altered cues.

### I.B.3. Experiment C

In Experiment C, blindfolds were used to investigate whether adaptation could occur in the absence of visual cues. Five blindfolded subjects performed 8 sessions of testing and training. Since subjects were blindfolded and could not accurately type responses, the identification response method was abandoned in favor of a pointing response: subjects were asked to turn their noses to point to the position of the click train after each presentation (subjects still had to face forward during each test stimulus or the trial was discarded). With the exception of the blindfold and the response method, Experiment C was identical to Experiment A.

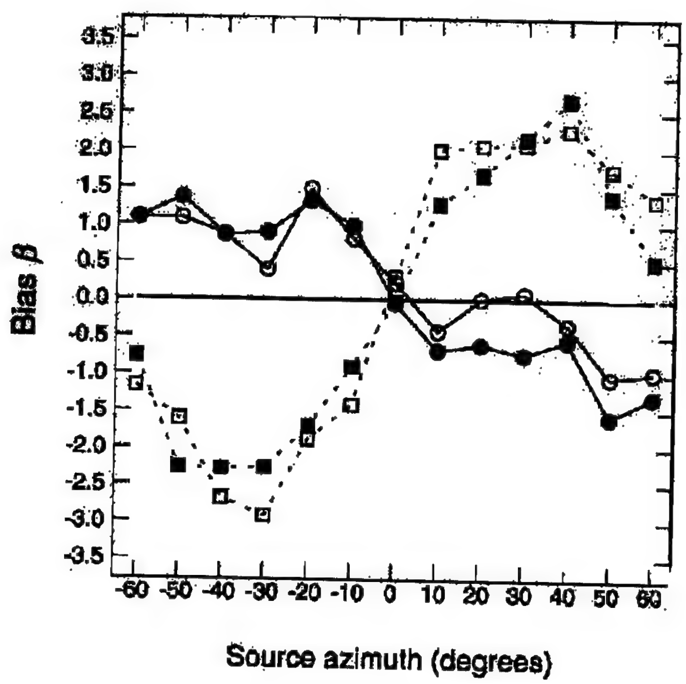


Fig. 7. Bias results for Experiment C. See Fig. 3 caption.

Bias results from Experiment C (shown in Fig. 7) are strikingly different from those of the previous experiments. No reduction in bias occurred with exposure, nor was there any negative after-effect.

In addition to the clear lack of adaptation with the experimental paradigm of Experiment C, other differences of note occurred. The edge effects seen in the previous experiments were much

less pronounced. Subjects were told verbally that only positions from -60 to +60 degrees in azimuth would be presented and were shown the possible source locations on the labelled light-arc prior to putting on the blindfolds at the start of each session, yet they still consistently turned outside the range of possible positions for altered cue sources at the edges of the azimuthal range. With normal cues, subjects showed a clear tendency to under-estimate the lateral position of the simulated sources, again in contrast to the previous experimental results. These differences are thought to be the result, at least in part, of the response method.

Resolution results for Experiment C are shown in Fig. 8. Resolution is somewhat enhanced in the central region (both before and after exposure) with altered cues, with the increase in resolution close to that seen in Experiment B. A slight decrease in resolution with exposure to the altered cues occurred in this experiment, but was not as pronounced as in Experiment B.

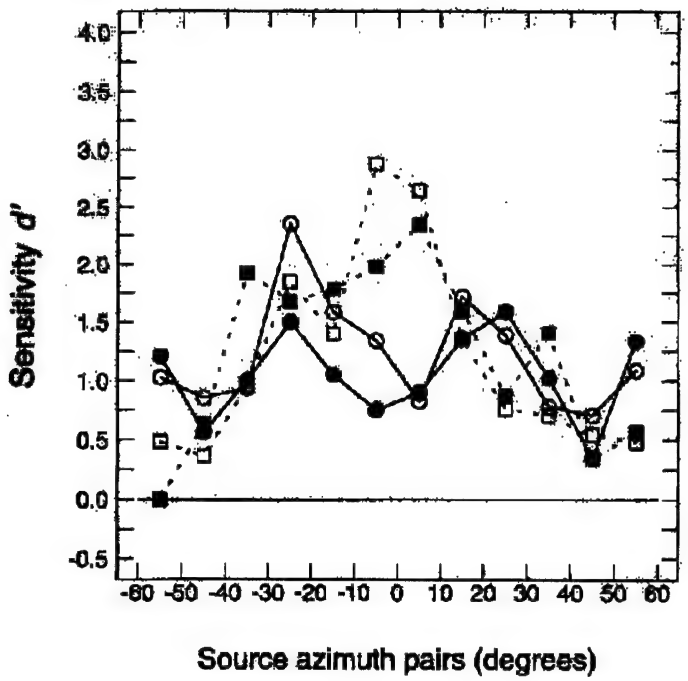


Fig. 8. Resolution results for Experiment C. See Fig. 3 caption.

### I.B.4. Experiment D

Experiment D was performed to test whether the lack of adaptation in Experiment C was the result of the altered response method or the lack of visual stimuli. The experimental paradigm used in Experiment D was identical to that of Experiment C, except that subjects were not blindfolded. The visual scene in the room was thus available to the subjects in this experiment, and subjects were exposed to correlated light/sound sources during training. Unfortunately, time limited the number of sessions performed by the four subjects who performed Experiment D: 3 of the subjects finished 2 identical sessions each, while the fourth finished 3 sessions.

Bias results from Experiment D are shown in Fig. 9. These data are clearly much noisier than any of the previous results. This is to be expected, since at least 4 times as many points were averaged in the previous results compared to those shown here. Although conclusions drawn from

the results of Experiment D are tentative at best, there does seem to be adaptation occurring for the data from the left side of the source range (from -60 to 0 degrees in azimuth).

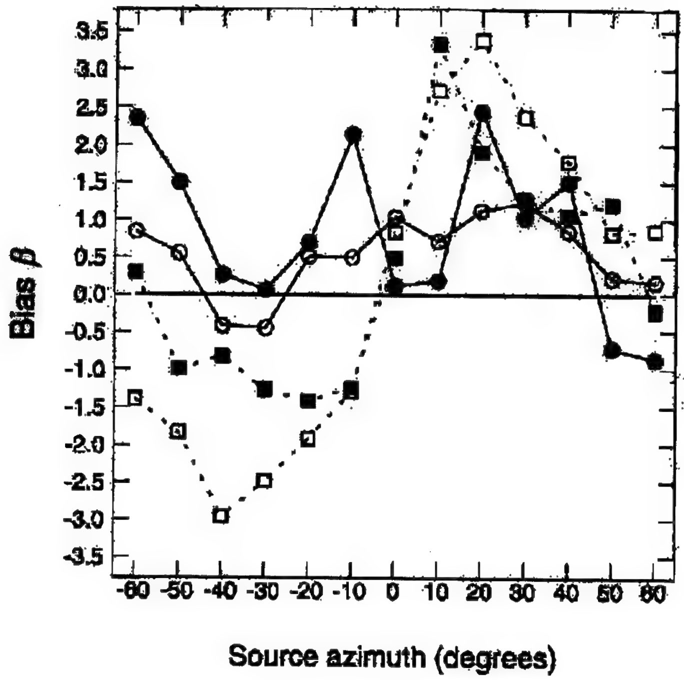


Fig. 9. Bias results for Experiment D. See Fig. 3 caption.

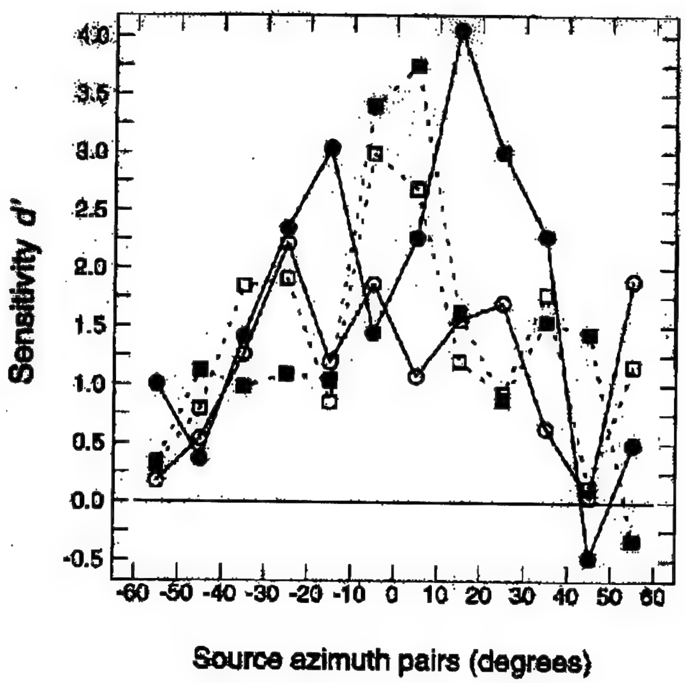


Fig. 10. Resolution results for Experiment D. See Fig. 3 caption.

These bias results are very similar to results from Experiments A and B. The usual strong

immediate effect is reduced in these data by nearly 50 percent with exposure to the altered cues, while a negative after-effect also occurs. On the whole, the results for the right side of the source range are not systematic. Examination of the raw responses for source positions to the right uncovered a large number of outliers in the responses for this half of the data. Given the small number of points averaged for the plots in Fig. 9, these outliers had a huge effect on the results for positions to the right of center, so that any effects which may have occurred were obscured by the noise.

Estimates of resolution for Experiment D are shown in Fig. 10. Again, the small amount of averaging for this experiment makes strong conclusions difficult. Resolution at the central two positions is elevated for both runs using altered cues; however, the random fluctuations in the normal run resolution data are larger than this resolution increase.

The results of Experiment D tentatively point to the blindfolding of subjects as the significant change in experimental paradigm between Experiments A and B and Experiment C. Time prevented detailed exploration of the dependence of adaptation on vision; however, the importance of vision to auditory spatial adaptation is not surprising. A large number of studies (Warren and Pick, 1970; Canon, 1970; Pick, Warren, and Hay, 1969; Jones and Kabanoff, 1975; Mastroianni, 1982; Platt and Warren, 1972; Ryan and Schehr, 1941) implicate vision as uniquely important in spatial perception.

### I.B.5. Experiment E

The first four experiments were done in a manner consistent with most previous work on adaptation, by using a training procedure that involves both the sensory and motor systems. In the psychophysical literature, training is often accomplished with correct-answer feedback, which is strictly cognitive in nature, and without motor involvement. To see if similar adaptation results could be obtained using general psychophysical procedures, Experiment E was performed without any active training runs, but with correct-answer feedback given after each trial by flashing the light at the correct location after the subjects entered his/her response.

In contrast to Experiments A and B, subjects never were given auditory and visual stimuli simultaneously, although visual stimuli were presented following auditory stimuli from the same location. Also, subjects did not experience localization cues involving the entire sensorimotor loop, since only testing runs (during which subjects faced forward during each presentation) were employed in Experiment E. As in Experiments A and B, subjects entered their responses on a keyboard rather than using the head-pointing response method. Three sources were present during every run (as in Experiment B). In order to make the exposure times similar to those of the previous experiments, 40 test runs of 26 trials each were used in Experiment E. Each session of 40 test runs lasted between an hour and an hour and a half. The order of the runs was

| | |
|---|---|
| 2 tests with normal cues | (1n-2n) |
| 8 tests with altered cues | (1a-8a) |
| - 5 minute break - | |
| 22 tests with altered cues | (9a-30a) |
| - 5 minute break - | |

8 tests with normal cues. (3n-10n)

In order to reduce variability, pairs of runs were analyzed together for the five subjects who performed 8 sessions of Experiment E. Thus, Runs 1n and 2n were averaged together across 8 sessions for each subject to give the normal cue baseline of performance; Runs 1a and 2a were combined to examine the immediate effect of the transformation; Runs 29a and 30a were averaged to examine the decrease in effect; and Runs 3n and 4n gave a measure of negative after-effect.

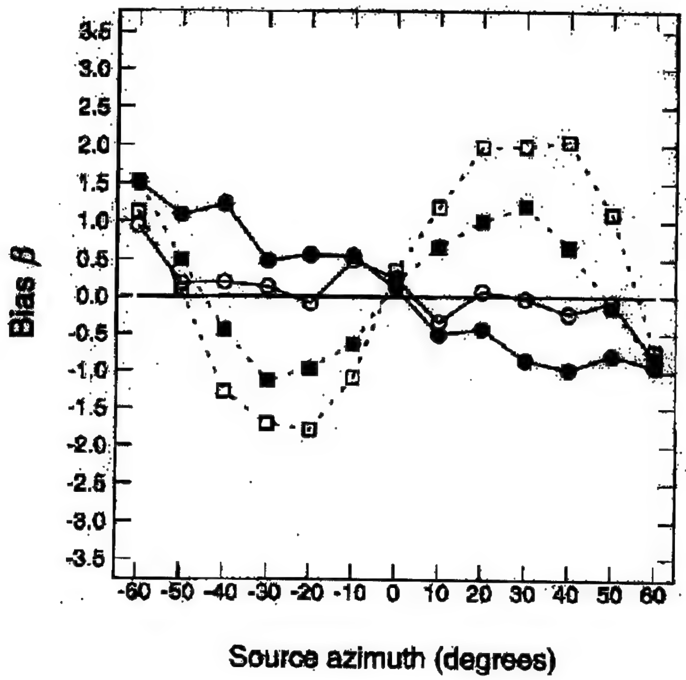


Fig. 11. Bias results for Experiment E. See Fig. 3 caption.

Bias results from Experiment E (shown in Fig. 11) closely resemble the results of Experiments A and B. An immediate effect is seen which follows predictions for the transformation and response method employed. The bias is reduced by about 30 percent with repeated exposure to the transformation (by correct-answer feedback in this case). When normal cues are tested following the altered cue tests, subjects show a strong negative after-effect.

Resolution results (shown in Fig. 12) are very similar to those of Experiment B. Resolution is enhanced in the first altered cue test for the center positions; however, this increase is reduced by the last altered cue tests. As in Experiment B (and unlike Experiment A), an ongoing music source was present from -90 degrees and a voice source from 180 degrees.

### I.B.6. Experiment F

The decrease in altered-cue resolution with time seen in all experiments but A, although in many cases of small magnitude, was surprising. Since peripheral resolution for the center positions was enhanced with the altered cues, it is reasonable to assume that the decrease in resolution over time must come from central mechanisms. Furthermore, if such were the case, then simplification of the task might eliminate the decrease over time.

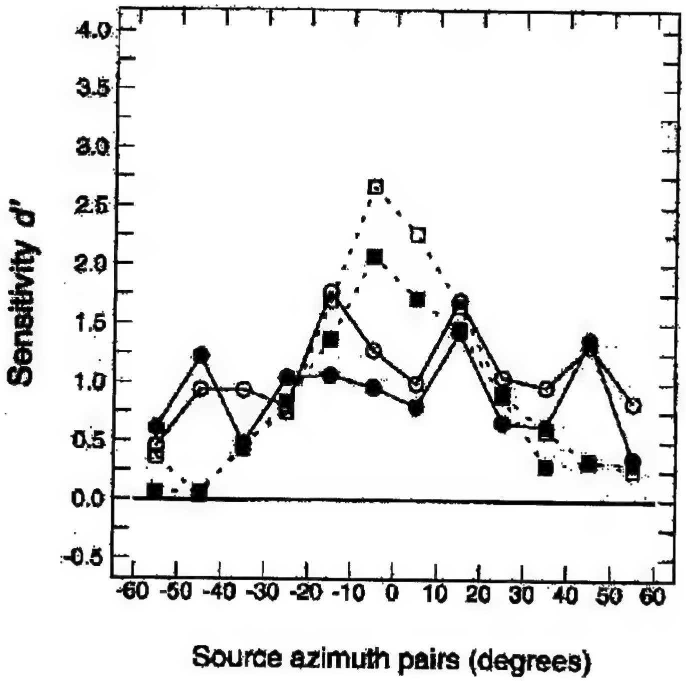


Fig. 12. Resolution results for Experiment E. See Fig. 3 caption.

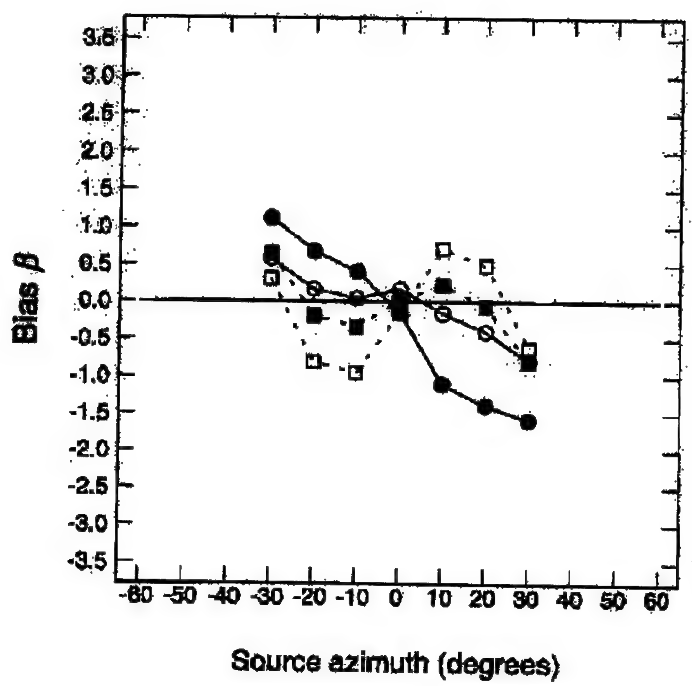


Fig. 13. Bias results for Experiment F. See Fig. 3 caption.

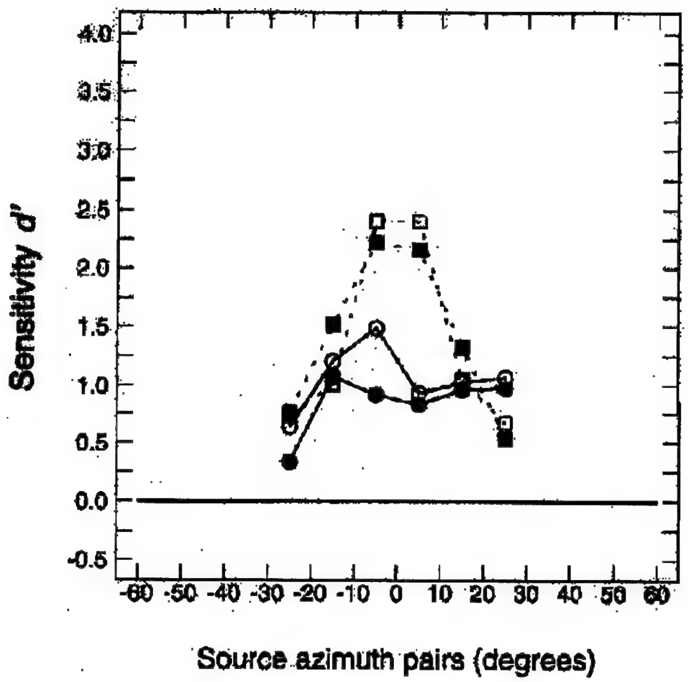


Fig. 14. Resolution results for Experiment F. See Fig. 3 caption.

With this in mind, Experiment F was performed using only the center seven locations. This change in the stimulus set simplified the task not only by decreasing the number of stimuli, but also by restricting the stimuli to a region where resolution always increased or remained unchanged (so that the resolution change was no longer non-monotonic). Experiment F was identical to Experiment E (with 2 continuous sources along with a target click train), except that only the seven center source positions were used.

Bias results for Experiment F, shown in Fig. 13, show the expected pattern of results. While the edge effect for Experiment F reduces the size of the immediate bias measured with the 7-alternative identification task, the bias is reduced by over 50 percent by the end of the altered-cue exposure period. The negative after-effect in Experiment F is at least as strong as was seen in previous experiments.

Resolution results are seen in Fig. 14. The results clearly show an increase in resolution for the center positions. Most importantly, resolution remains enhanced throughout the altered-cue exposure time.

### I.B.7. Hand-Pointing Experiments

A number of the previous experiments investigating auditory adaptation (e.g., Mikaelian and associates, Freedman and associates, and Kalil) employed paradigms in which a sound source was held in the hand of the subject. In these experiments, cues were altered with a pseudophone while the subject made pointing responses with the hand holding one source to match the position of a target source. In order to determine whether shifting the paradigm in this manner would alter our results, another testing paradigm was developed which used the Convolvotron in conjunction with

a tracker worn on the hand. In these experiments, subjects were seated with their heads held stationary in a head rest. A target source was presented at one of ten positions around the subject, and the subject was asked to make a ballistic pointing movement with his right hand to match the azimuthal position of the target. When the hand reached the end of its trajectory, a source simulated at the hand was turned on, and subjects heard the extent of their pointing error. This experimental paradigm was perfected in a series of pilot tests, and we are now ready to begin more formal tests.

## I.C. Modeling Efforts

The most important accomplishment to date has been the development of a model capable of describing the experimental results for nearly all of the experiments performed so far. In particular, the model predicts the changes in both bias and resolution as subjects adapt to supernormal localization cues.

This model is significant for both sensorimotor adaptation and psychoacoustics. Its importance in the field of sensorimotor adaptation arises from two main facts. First, the model can make quantitative predictions of most aspects of performance over the time course of the adaptation process. Not only are most previous adaptation models qualitative rather than quantitative, but other models are generally restricted to describing only how mean performance changes over time; no attention is given to how resolution varies over time. As such, the current model is more powerful than any existing model of sensorimotor performance found in the literature. Second, the necessary assumptions of the model have important implications for sensorimotor adaptation in general.

The model assumes that 1) subjects cannot adapt to nonlinear transformations of auditory localization cues, but instead adapt to a linear approximation of the imposed nonlinear transformation, and 2) given the linear constraint on adaptation, performance asymptotes to levels very close to the ideal levels achievable for the imposed nonlinear transformation. The fact that these assumptions allow the model to describe the results from the current experiments leads to a number of important questions about other sensorimotor adaptation studies. For instance, is it a general principle that humans cannot adapt to nonlinearities in sensorimotor rearrangement? Can it be shown that the failure of humans to completely adapt to other types of rearrangement is really a by-product of their inability to adapt to nonlinear transformations? How universal are the principles found in the current model? These questions are of great importance both for scientific understanding of sensorimotor adaptation, and for practical applications in which human subjects must adapt to sensorimotor rearrangements.

The importance of the model in the field of psychoacoustics arises from its ability to describe how performance evolves over time when subjects are provided with specific feedback or training that pushes them to interpret physical stimuli in new ways. No existing psychophysical models provide insight into how aspects of performance might change over time.

The current model incorporates ideas from the fields of sensorimotor adaptation and psychoacoustics to lead to a model that significantly adds to both areas. Whereas previous models

of sensorimotor adaptation provided only qualitative descriptions of how one aspect of performance changed over time, previous psychophysical models provided detailed quantitative predictions of all aspects of performance at one point in time. The current model of adaptation to supernormal auditory localization cues provides quantitative predictions of all aspects of performance over time.

In the model, it was assumed that the range of stimuli attended to at any given time (determined by the slope relating mean response to acoustic stimulus) determined the amount of noise in the human perceptual system. For larger ranges of stimuli, noise increased, whereas for smaller stimulus ranges, it decreased. In the model, as subjects adapted to supernormal cues, the range of stimulus values to which the subject attended also increased. This increase in range made the model predict a decrease in resolution over time for the same physical stimuli.

## I.D. Computations Concerning the Use of Frequency-Scaling to Simulate an Enlarged Head

In examining ways in which supernormal localization cues could be produced, the idea of generating HRTFs from larger-than-normal heads was considered. Large-head HRTFs could be tested with equipment and experimental paradigms similar to those used in the previous experiments, once the large-head HRTFs were produced. One way of generating large-head HRTFs would be to build a physical model of a larger-than-normal head, and to empirically measure the resultant cues. This method is not only very time consuming, but also inflexible, since for every new head-size to be tested, the whole procedure would have to be repeated.

An alternate approach would derive large-head HRTFs from empirically measured, normal HRTFs. One method for doing this (already discussed in Sec. B-4-a-i) is to use frequency scaling. In anticipation of employing this method, the theoretical effects of frequency-scaling HRTFs to approximate a larger than normal head were investigated and reported in Rabinowitz, Maxwell, Shao, and Wei (1993). In this work, it was shown that frequency-scaling normal HRTFs will produce results very similar to HRTFs from larger than normal heads, provided the sources to be simulated are relatively far from the listener.

## I.E. Distance Coding

As indicated previously, neither distance nor elevation are well perceived naturally. Furthermore, very little effort has been made to explore the extent to which perception of these variables could be substantially improved for use in acoustic displays by means of artificial coding. In fact, and as pointed out in Sec. B-4-b, it is only recently that individuals concerned with virtual acoustic displays have begun to perform experimental work related to the simulation of natural distance coding.

We have completed an initial study concerned with subjects' abilities to identify various filter transfer characteristics (Brungart, 1994), using characteristics that we might use for coding distance (and/or elevation). This work differs from past work performed at a variety of laboratories, including our own, on the perception of spectral shape (Green, 1988; Durlach,

Braida, and Ito, 1986; Farrar, Reed, Ito, Durlach, Delhorne, Zurek, and Braida, 1987) in that we are interested in selecting shapes and experimental paradigms related to the coding of distance and/or elevation rather than in modeling of cross-frequency intensity comparisons. In particular, in most cases our attention is focused on (1) transfer characteristics that are related (at least loosely) to those naturally encountered and that do not interfere too seriously with the perception of the transmitted signal (i.e., with the "message") and (2) the task of identification rather than discrimination. Our initial series of experiments examined absolute identification (AI) performance using the 3-dimensional set of "single-echo" filters given by (and specified previously by Eqs. B-12 in Sec. B):

$$\left|S_{A,m,\tau}(\omega)\right|^2 = A^2\left[1 + m^2 + 2m\cos(\omega\tau)\right]$$
$$\text{phase}\left[S_{A,m,\tau}(\omega)\right] = \tan^{-1}\left[\frac{-m\sin(\omega\tau)}{1 + m\cos(\omega\tau)}\right]. \qquad \text{(B-12)}$$

We have investigated the ability of listeners to perceive information encoded as the strength and delay of a single echo of the source. Although the work focused on how much information listeners can extract when distance-like cues are presented rather than on the perception of distance per se, it is a first step toward developing simple but reliable distance cues for a virtual-environment system.

The most important result from these experiments was that the amount of information transfer (IT) was startingly small. Whereas many unidimensional stimulus sets lead to an IT of 2 to 3 bits (and two-dimensional stimulus sets to an IT of 3-5 bits), in these experiments the value of IT obtained fell in the range of 0.2 - 2 bits. In general, these small values of IT appeared to result from two factors: (1) large JNDs in the variables m and $\tau$ and (2) lack of perceptual independence between m and $\tau$ (i.e., discrimination or identification of one variable increased substantially when the value of the other variable was randomized). Thus, it appears that encoding distance solely by means of the single-echo modulation parameters m and $\tau$ cannot lead to good distance resolution.

# II. Personnel Supported

The personnel supported by and/or associated with this project are Principal Investigators Nat Durlach and Dick Held; post-doctoral associate Barbara Shinn-Cunningham; research specialist Lorraine Delhorne; and graduate students Greg Lin, Kinu Masaki, and John Park.

# III. Interactions/Transitions

## III.A. Publications

Durlach, N. I. (1991). "Auditory Localization in Teleoperator and Virtual Environment Systems: Ideas, Issues, and Problems," Perception, 20, 543-554.

Durlach, N. I., Rigopulos, A., Pang, X. D., Woods, W. S., Kulkarni, A., Colburn, H. S., and Wenzel, E. M. (1992). "On the externalization of auditory images," Presence, 1, 251-257.

Durlach, N. I., Shinn-Cunningham, B. G., & Held, R. M. (1993). Supernormal auditory

localization. I. General background. Presence, 2(2), 89-103.

Rabinowitz, W. R., Maxwell, J., Shao, Y., & Wei, M. (1993). Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere. Presence, 2(2), 125-129.

Shinn-Cunningham, B. G., and Durlach, N. I. (1994). "Defining and redefining limits on human performance in auditory spatial displays," in Auditory Display, Ed. Greg Kramer and S. Smith. Santa Fe: Santa Fe Institute.

Shinn-Cunningham, B. G., Zurek, P. M., Durlach, N. I., and Clifton, R. K. (1995). "Cross-frequency interactions in the precedence effect," J. Acoust. Soc. Am., 98(1), 164-171.

Shinn-Cunningham, B. G., Lehnert, H., Kramer, G., Wenzel, E. M., and Durlach, N. I. (1996). "Auditory Displays," in Spatial and Binaural Hearing, Eds. R. Gilkey and T. Anderson. New York: Erlbaum (in press).

Shinn-Cunningham, B. G., and Kulkarni, A. (1996). "Applications of Virtual Auditory Space," in Virtual Auditory Space, Ed. S. Carlile, Landes Publishing Company. New York (in press).

## III.B. Talks

Durlach, N. I. (1991). "Sensing and Displaying Acoustic Information," ILP Symposium on Telerobotics, MIT, Oct. 29-30, 1993.

Durlach, N. I. (1991). "Super Auditory Localization for Improved Human-Machine Interfaces," DOD User-Computer Interaction Technical Group, San Antonio, TX, Nov. 5, 1991.

Durlach, N. I., Held, R. M., and Shinn-Cunningham, B. G. (1992). "Super Auditory Localization Displays," Society for Information Display International Symposium Digest of Technical Papers, vol. XXIII, 98-101.

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. (1992). "Adaptation to transformed auditory localization cues in a hybrid real/virtual environment," J. Acoust. Soc. Am., 92, 2334.

Shinn-Cunningham, B. G. (1993). "Auditory virtual environments," talk presented at the M.I.T. Workshop on Space Life Sciences and Virtual Reality, Endicott House, Dedham, MA, 6 January 1993.

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. (1993). "Super Auditory Localization for improved human-machine interface," talk presented at the AFOSR Review of Research in Hearing, Fairborn, OH, June 1993.

Shinn-Cunningham, G. B. (1993). "Auditory displays and localization," talk presented at the Conference on Binaural and Spatial Hearing, sponsored by the AFOSR and Armstrong Laboratory, Wright Patterson AFB, September 9-12, 1993.

Shinn-Cunningham, B. G., Delhorne, L. I., Durlach, N. I., and Held, R. (1994). "Adaptation to supernormal auditory localization cues as a function of rearrangement strength," J. Acoust. Soc. Am., 95, 2896.

Shinn-Cunningham, B. G. (1995). "A dynamic, psychophysical model of adaptation in

localization experiments.," J. Acoust. Soc. Am., 97(5), 3411.

### III.C. Meetings

Additional work connected with this grant has involved participation in meetings at government agencies (e.g., NASA and ONR) and participation in meetings of the Acoustical Society of America.

## IV. New Discoveries, Inventions, or Patent Disclosures

An invention disclosure has been submitted to M.I.T.'s Office of Technology Licensing for the work on an inertial tracking system. A patent may ensue for this tracker, which was supported both by this project and NASA contract NCC 2-771.

# TRANSOM Progress Report
# January -February 1996

*David Zeltzer*
Sensory Communication Group
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge MA 02139

## TRANSOM Sensorimotor Involvement Experiments

During the month of January, all the basic software and hardware set-up required to start the 1 degree-of-freedom Dynamic Position and Orientation Control experiments was completed. The basic dynamics simulation (including dynamic models for the ROV, the "target", and a predictor) was completed. In addition, the ability to run experimental scripts and automatically capture data was implemented.

During the month of February, the experimental set-up and pilot runs for the initial 1 degree-of-freedom Dynamic Position and Orientation Control experiments were completed. Due to improvements made last month to the experimental software, the planned schedule was accelerated. Scenarios (i.e., training and transfer conditions) were developed for the following situations:

1) Discrete Target Jumps and No External Influences on the ROV;
2) Continuous, Dynamically-Lawful, Target Moves and No External Influences on the ROV; and
3) Stationary Target and ROV Influenced by Current.

Seven subjects were started on these experiments during the month. All subjects completed the first scenario and are in the process of completing the other scenarios.

An initial concept design for an ROV Path Awareness environment has been completed. Work has started on the virtual environment to be used by experiments. The ability to precisely locate reference objects within the VE has been completed.

During March, we expect to refine experimental design and to complete a generic control device interface to the VE so that various viewpoint control methods can be tested.

## Baseline ROV simulation

During the month of January, the 3D user interface was improved so that each virtual joystick rotates about a single point, as one would expect in a regular joystick. In addition, the joysticks were modified to appear to be spring loaded and centering.

A C++ "wrapper" was created around the dynamics code provided by Imetrix, and this uses a Scheme interface and a Header2Scheme module for interfacing with C++. Most of the libraries were converted to Dynamic Shared Object (DSO) libraries to increase linking time without significant loss at startup time. Finally, for increased

integration with the Scheme system, the C++ simulation code was rewritten as Scheme code.

Utilities were implemented to make it easier to modify environmental characteristics, such as fog-type and visibility, etc.

In February, the collision detection library was experimented with , and it is now very close to being integrated into the Scheme system. The 3D interface was modified to include an ROV-centered viewing position, as well as the previously programmed "bird's eye view".

A new scenario was created for the ROV, in which component pieces of an F-16 model are scattered around the ocean floor. A B-spline surface was created to model the undersea terrain, and an “ocean floor” texture was texture-mapped on this spline surface.

Finally, per the scenario, functionality was implemented to allow the developer to take "pictures" of parts of the crashed aircraft, which can be displayed at the end of a session.

During March, we expect to continue to elaborate the models, adding collision detection, and developing a model of the virtual control console, so that the ITS can access data regarding user control input to the ROV. Additional controls will be added to the GUI console to control floodlights on the ROV, and to control the tilt of the simulated video camera.